Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference "Dialogue 2025"

April 23–25, 2025

Operator words versus MT systems and LLMs[1]

Zimmerling A. V.
Pushkin State Russian Language Institute;
Institute of Linguistics, Russian Academy of Science

Baiuk A. M.
Russian State University
for the Humanities

Abstract

This paper investigates the meanings of two classes of Russian discourse words, defined by modal operators VER and AFF, and examines their translation equivalents in nine target languages using machine translation (MT) systems, large language models (LLMs), and human expert translations. VER (verification) indicates confirmation of a hypothesis, while AFF (affirmation) expresses a strengthened belief in a hypothesis. The study uses a set of 17 Russian discourse words and evaluates their translations in English, German, Danish, Swedish, Icelandic, Ukrainian, Bulgarian, Ossetic, and Arabic across 85 test sentences. The primary goal was to test the universality of the VER and AFF distinction, hypothesizing that these classes remain distinct across languages despite the lack of direct one-to-one translation equivalents. The study assumed that VER and AFF operators, corresponding to DE RE and DE DICTO attitudes respectively, differ fundamentally in semantics and distributional behavior. The study confirms the semantic and distributional independence of VER and AFF operators, supporting their universality. LLMs, despite not being specialized for MT tasks, showed remarkable adaptability and context awareness compared to traditional MT systems. The findings highlight the potential of LLMs in nuanced translation tasks and underscore the complexity of translating modal discourse words. Future work will explore custom models and further refine evaluation metrics for translation accuracy.

Keywords: discourse words, operators, machine translation systems, large language models
DOI: 10.28995/2075-7182-2025-23-XX-XX

Модальные операторы в автоматических системах перевода и больших языковых моделях

Циммерлинг А. В.
Государственный институт русского языка имени А. С. Пушкина; Институт языкознания РАН

Баюк А. М.
Российский государственный гуманитарный университет

Аннотация

Обсуждаются значения двух классов русских дискурсивных слов, определяемых модальными операторами VER и AFF, и рассматриваются их переводные эквиваленты на девяти целевых языках с использованием машинных переводчиков, больших языковых моделей и переводов, выполненных экспертами. Оператор VER (верификация) обозначает подтверждение гипотезы, в то время как AFF (аффирмация) выражает усиленную уверенность в гипотезе. В исследовании использован набор из 17 русских дискурсивных слов и оцениваются их переводы на английский, немецкий, датский, шведский, исландский, украинский, болгарский, осетинский и арабский языки на основе 85 тестовых предложений. Основной целью было проверить универсальность различия между VER и AFF на основе гипотезы о том, что различение этих классов сохраняется во всех языках, несмотря на отсутствие пословных эквивалентов. Предполагалось, что операторы VER и AFF, соответствующие отношениям DE RE и DE DICTO, фундаментально отличаются по семантике и дистрибутивным свойствам. Исследование подтверждает семантическую и дистрибутивную специфику операторов VER и AFF. Большие языковые модели продемонстрировали большую адаптивность

[1] This paper was written with support from the RSCI project № 22-18-00528 "Clausal connectives in sentence and discourse: Semantics and grammaticalization paths" realized at the Institute of Linguistics. We are grateful to the anonymous reviewers for the critical comments and to Svetlana Polukhina for the help with assessing the data. All disclaimers follow.

и способность учитывать контекст по сравнению с традиционными системами машинного перевода. Результаты подчеркивают потенциал больших языковых моделей в решении задач перевода и акцентируют внимание на сложности перевода модальных дискурсивных слов. В будущих исследованиях планируется разработка новых моделей и доработка метрик оценки точности перевода.

Ключевые слова: дискурсивные слова, модальные операторы, машинный перевод, большие языковые модели

1 Introductory remarks

In this paper, we check the meanings of two groups of Russian discourse words and their translation equivalents provided by MT systems, LLMs and human experts. We selected 17 Russian discourse words and multiword expressions presumably containing two kinds of modal operators VER and AFF. Their values are defined as follows:

(i) VER (*p*): *X* confirms the hypothesis *p* about the outer world.
(ii) AFF (*p*): *X* states that his belief that *p* is strengthened.

The scenario VER entails that the speaker considered the hypotheses *p* and ~ *p* and verified *p* at the moment *t* [33]. The scenario AFF entails that the speaker was originally biased towards *p* at the moment *t-1* and is still biased towards *p* at the moment *t*, i.e. the moment of speech. Russian has a large amount of discourse words that have been discussed in detail [3; 10; 16], cf. also [24; 18: 122; 35] on the language-specific discourse word *pravda*, [11] on *razumeetsya*, *konechno*, *samo soboj*, *estestvenno* and [23] on the particle *taki*. For this study, we assumed that Russian has two classes of VER and AFF words that can have external correspondences in other languages. Class 1 contains 7 standard elements expressing VER. 6 of them always take over the phrasal accent, while *taki* is an enclitic.

(iii) Class 1 words: VER (*p*): *dejstvitel'no*, *real'no*, *(i) pravda*, *na samom dele*, *v samom dele*, *taki, i vpryam'*.

Class 2 contains 9 standard elements expressing AFF.

(iv) Class 2 words: AFF (*p*): *razumeetsya*, *konechno*, *samo soboj*, *estestvenno*, *ponyatno*, *yavno, ya tak i znal, tochno, opredelenno.*

The 17th item, Rus. *konkretno*, is a recent colloquial word that presumably patterns with Class 1. We however left the final verdict to the translators and MT systems depending on their ability to reconstruct VER or AFF equivalents in the target languages.

For the translation task, we offered 5 short Russian sentences containing a slot for a VER | AFF operator:

Fig. 1. Stimuli sentences (Russian)

1) Vasya VER | AFF durak 'Vasya is VER | AFF a fool'.
2) Vasya VER | AFF ne pridet 'Vasya VER | AFF won't come'.
3) Vasya VER | AFF ne prishel 'Vasya VER | AFF didn't come'
4) Vasya VER | AFF oshibsya 'Vasya VER | AFF made a mistake'.
5) Vasya VER | AFF ne sobiralsya prixodit' 'Vasya VER | AFF wasn't going to come'.

The number of the stimuli totals 17 x 5 = 85. The stimuli were translated into 9 languages — English, German, Danish, Swedish, Icelandic, Ukrainian, Bulgarian, Ossetic and Arabic — by 1) the human expert[2]; 2) MT systems (Google, Yandex and the pre-trained Google model); 3) LLMs (ChatGPT 4o and Gemini 1.5). The experts produced 765 target sentences (85 x 9) altogether. The same amount of target sentences was produced by each MT system and LLM. In addition, we checked the responses of MT systems and LLMs by giving them both Russian stimuli sentences with the bare operator word

[2] The group of experts (6 women and 3 men, from 30 to 88 years) included one bilingual person and two L1 speakers of the target language with a near-native level of Russian. The rest were L1 speakers of Russian with a near-native or highly proficient level of the target L2 language. All experts but one had a linguistic background, were engaged in the teaching of foreign languages and had a translation experience.

pravda^{VER} and with the added proclitic *i* (*i pravda*^{VER}). The experts took the equivalence *pravda*^{VER} ≅ *i pravda*^{VER} for granted, but the MT systems and LLM reacted to these stimuli differently, so we got extra 5 target sentences from each MT system and LLM.

The goal of our study is to check the hypothesis the distinction of Class 1 words containing VER versus Class 2 words containing AFF is universal despite possible asymmetries in the number of Class 1 and Class 2 elements across the world's languages and the lack of exact word-to-word equivalents of the tested Russian discourse words[3]. The translation was valued as semantically correct, if the source sentence containing VER or AFF was rendered by the target sentence with the same operator. The zero hypothesis was that Class 1 and Class 2 do not intersect in any language, so that no discourse word that is a standard part of the Class 1 lexicon patterns with to Class 2 in any of its uses, and vice versa.

2 Basic linguistic intuitions

Verification and affirmation are different meanings, despite they are occasionally conflated in the description of modals, cf. [21: 81; 19: 78; 2: 33] and impressionistic labels like 'enhanced indicativity' used for verification markers in [29: 299]. Informally, verification of *p* does not imply that *X* was originally biased towards *p,* and the fact that *X* preserved his belief that *p*, does not guarantee that *p* is true. More specific arguments for treating VER and AFF as independent operators come from the distribution of verification and affirmation markers in the world's languages. The meaning of verification can be encoded by the intonation [28: 82; 17], moreover, verification words normally get the phrasal accent, cf. *dejstvitel'no* 'really', *pravda* 'true that' that are always accented in Russian [27], while affirmative markers like Rus. *estestvenno* 'naturally', *razumeetsya* 'certainly' can be both accented and deaccented [36]. Further linguistic tests involve the asymmetry of VER and AFF words in the so-called indirect contexts, i.e. utterances about the beliefs of other people. Since VER words freely apply to all possible worlds given that *p* is true in some accessible possible world *W*, cf. *If X* ***really*** *were a good boy, he would buy his girl a gift*, they are licensed both in the counterfactual and in the indirect contexts, cf. (1a-b).

(1) Russian

a. Esli by Vasya **dejstvitel'no** opozdal, ego by oshtrafovali.

'If Vasya **really** came late, they would fine him.'

b. Petya (**dejstvitel'no**) dumaet, chto Vasya **dejstvitel'no** opozdal.

'Pete (**indeed**) believes that Vasya **indeed** came late.'

VER words normally combine with the irrealis markers and can be used in conditional clauses in the subjunctive mood, cf. [33] on the Old Russian verification particle TI_1 'indeed'. AFF words however do not combine with the irrealis markers and are blocked in the indirect contexts, since the speaker cannot project his own belief that *p* to other people, cf. (2a-b)[4].

---

[3] For modal particles in some of the target languages see [1; 15; 2].

[4] An anonymous reviewer points out that AFF items can be analyzed as embedded to a different depth, cf. Krifka's and Frey's accounts for German and English [14; 7]. This holds for the Russian AFF words as well, cf. the context provided by the reviewer, where *konechno* 'certainly' cannot be replaced by *estestvenno* 'naturally': *Ya*^{EGO} *eshche dumal, chto, konechno*^{EGO}*, ego i menya ravnyat' glupo* 'I^{EGO} thought that it was, of course^{EGO}, stupid to compare him and me', $^{??}$*dumal, chto, estestvenno*^{S}*, ego i menya ravnyat' glupo*. This contrast can prima facie be explained by the fact that *konechno* is an egocentric word projecting the point of view of the speaker / propositional subject [22: 311], while *estestvenno* and *naturally* introduce an assessment ('judgment', in Krifka's terms) shared by the speaker and the addressee. However, the classification of AFF words is orthogonal for our paper: we do not claim that Class 2 items containing AFF are absolute synonyms and replace each other in all root and embedded contexts. Neither do we claim this for Class 1 items containing VER. What we check, is the ability of translators to extract the operators VER and AFF from the corresponding word classes correctly.

(2) Russian

a. *Esli by Vasya **estestvenno** opozdal, ego by oshtrafovali.

Int. 'If Vasya **naturally** came late, they would fine him.'

b. *Petya (**estestvenno**) dumaet, chto Vasya **estestvenno** opozdal.

Int. 'Pete (**naturally)** believes that Vasya **naturally** came late.'

Class 1 items containing VER can be straightforwardly identified with *de re* modals, i.e. the attitude to describe the real or accessible possible world as it is. Class 2 items containing AFF can arguably be linked with *de dicto* modals, i.e. the attitude to interpret the text <on the model set of possible worlds> in the form as it is expressed. The proof of these claims is beyond the reach of this paper. We just assume here that the distinction of Class 1 versus Class 2 holds on the level of logical semantics and is universal[5]. In the following, we conventionally label Class 1 items 'DE RE words' and Class 2 items 'DE DICTO words'[6]. Most Russian DE RE and DE DICTO words are homonymic to standard adverbs, nouns, cf. *pravda* 'truth', adjectives, cf. *ponyatno* 'it is clear' and even clauses, cf. *Ya tak i znal* 'I knew it'. Therefore, the translation of DE RE and DE DICTO words amounts to a triple task: 1) recognizing the VER | AFF insertion as a discourse item in the position where it occurs in the source sentence; 2) extracting the correct operator from the discourse item; 3) providing a suitable equivalent from the set of DE RE or DE DICTO words in the target language. This is schematically shown in (v); the symbol W^{VER} stands for the set of all discourse elements containing VER, 'SL' stands for the source language, and 'TL' for the target language.

(v) $w_i \in \{_{SL}\, w_1, w_2 \ldots w_n\}^{DE\ RE} \Rightarrow W^{VER} \Rightarrow w_j \in \{_{TL}\, w_1, w_2 \ldots w_n\}^{DE\ RE}$

3 Linguistic models and the design of the experiment

In this section, we comment on the chosen linguistic model and its impact on the design of our experiment. We assume that verification is a universal meaning that is encoded lexically by the words containing VER, by prosody, or by both cues [27]. We also adopt the hypothesis that Russian, which is the source language in the experiment, has a class of VER words that have a different distribution compared to AFF words [36]. Provided that the meaning of VER is universal, this gives a ground to check whether the contrast between VER and AFF words is preserved in translation. Since operator semantics represents a more abstract layer of meaning than contextually-determined uses, our approach is complementary with the previous descriptions of Russian discourse words [3; 10; 11; 16] that focus on their cognitive profiles. Krifka's layered model of speech acts distinguishes two kinds of abstract objects called 'commitments' and 'judgments' and identifies them with syntactic projections [13; 14]. This model describes the semantics-to-syntax interface but leaves no space for VER words[7] and does not suit our experiment.

An anonymous reviewer argues that artificial stimuli are not really valid as MT systems and LLMs have been pre-trained on rich sentences, but our stimuli are grammatically correct, and their shortness shouldn't affect anything, besides, training data can't contain only long (rich) ones. On the other side, prototype sentences with similar contexts may allow to evaluate the ability of neural models to distinguish the operator words themselves. The linguistic problem is that the set of our short lab sentences invites the translator to reconstruct the contexts, where these sentences are appropriate:

[5] The DE RE / DE DICTO distinction can be interpreted differently in logical studies. The analysis of verification predicates as DE RE modals is close to the approach dubbed 'metaphysical' by Michael Nelson [20].
[6] An anonymous reviewer objects that our categories "DE RE words" and "DE DICTO words" duplicate the labels VER and AFF. We prefer to withhold the double notation since the DE RE / DE DICTO terminology applies to sentences and propositional attitudes, while VER and AFF labels can apply to the sentence components and to the storage of operator words in the lexicon.
[7] This issue is acknowledged by Krifka himself, who admits that German VER words *tatsächlich* and *in der Tat* 'indeed', 'in fact' are not commitment modifiers but elements invoking 'the contrast between the proposition and its negation, similar to cases of verum focus' [14: 146].

otherwise a considerable part of the stimuli is redundant. Another anonymous reviewer suggests that we drop our 17 ‘lab stumps’ and import a much richer collection of genuine examples from the parallel corpora. This is unfortunately not realistic since the diversity of the reconstructed contexts for our lab stimuli with 17 discourse words far exceeds the limits of the parallel corpora even for well-documented languages. The same reviewer asks what MT and LLM contribute compared to human translation in our experiment. To begin with, neural models are based on statistical language modeling and pre-trained on huge amounts of natural language data. Therefore, they may emulate human perception of operator words and cross-lingual connections between them. Finally, the quality of the MT and LLM translation of discourse words may reflect the gaps in the training data, and we check this in the experiment.

4 Translation issues

There are two extremes in translation studies. One group of scholars basing on monolingual dictionaries and lexicographical definitions capitalize the role of language-specific elements and their functional equivalents in the parallel corpora [30; 23; 5]. Discourse words are language-specific, since they have diverse correspondences in corpora, but such correspondences can be registered in super-corpus data bases [31; 32] and eventually added to monolingual dictionaries [6]. Other authors deny the existence of translation equivalents on the word-level and argue that discourse words lack their own meanings [19]. The last claim echoes the well-known theory by Victor V. Vinogradov that only content but not function words have lexical meanings. We are not going the mediate this dispute here, cf. the opposite statements on the need to account for function words in the lexicon [9; 34]. However, we would like to add an extra dimension to translation of operator words, namely, the distinction of *horizontal* and *vertical* translation. While horizontal translation involves an attempt to use one and the same target equivalent in all contexts, vertical translation is context-bound: a group of semantically close words tends to get similar or identical targets in the same context, but one source word gets different targets in different contexts. The vertical strategy for tackling Russian DE DICTO words is schematically shown in fig. 2: the symbols like ‘*b’ stand for the slots where the translator failed to select the target predicted by the vertical strategy.

	Context 1	Context 2	Context 3	Context 4	Context 5
razumeetsya ^{AFF}	a	b	c	d	e
konechno ^{AFF}	a	b	c	*c	e
samo soboj ^{AFF}	a	b	c	d	*b
estestvenno ^{AFF}	a	b	c	*c	*f
ponyatno ^{AFF}	a	b	c	d	e

Table. 1: Translation grid for 5 Russian DE DICTO words (vertical translation)

The vertical strategy is based on the translator’s capacity and will to look up or reconstruct the broad context, which is of double importance in the case of short lab sentences in our study, where neither the human experts nor the MT systems had access to prosody and coherent text fragments. We assume that this factor is relevant for the expected contrast between expert and MT translation as well as for the choice of the horizontal versus vertical strategy by MT systems and LLMs.

The recognition of VER and AFF brings about a more specific problem, notably, the overtness of the operator in the target language. Since verification can be encoded solely by intonation across the world’s languages, we accepted the omission of the lexical VER operator ($w_{SL}{}^{VER} \Rightarrow \varnothing^{VER}$) as a legal move. Meanwhile, affirmation is always encoded lexically, therefore the dropping of the AFF operator is illegal. E. g., both Google MT and Yandex MT translate the source sentence with *Ya tak i znal* ^{AFF} into German without the modal particle *ja* prescribed by our expert, therefore we considered the translation (3) a mistake.

(3) Russian ⇒ German

> [14.2.] Rus. **Ya tak i znal**[AFF], chto Vasya ne pridet ⇒ Ger. **Ich wusste Ø**, dass Vasya nicht kommen würde. (Google MT & Yandex MT).

5 5Technical aspects

In this work, several MT options were used: a) MT systems Google[8] and Yandex[9]; b) Large Language Models (LLMs) ChatGPT 4o [25; 4] and Gemini 1.5 [8]. We had also been planning to use a MT model based on Transformer architecture [26] (namely, T5 or Text-to-Text Transfer Transformer), but there are few pre-trained options with all the languages we need, and there were limits to our computational resources so we weren't able to train a custom model, although we may do so for the future research. We have conducted several experiments with a one-to-many languages MT model[10], but it translates from English to other languages, so it would be questionable to compare the results. It should be noticed, though, that the use of an open pre-trained MT model allows one to get token embeddings for further investigations, visualisation included.

Modern MT systems such as Google or Yandex Translate are also Transformer-based[11], but at least for Google Translate there are reports[12] that there are language pairs it doesn't translate directly, but with the use of an intermediate language (most often, English). We have chosen Google and Yandex Translate because of their popularity.

As for the LLMs, it is known that they weren't designed specifically for the task of MT, but are good at it, as research shows [12]; ChatGPT 4o is currently one of the leading LLMs in known benchmarks[13], whereas Gemini 1.5 is a model with a much smaller parameter size (200 billion vs 1 trillion), and this was one of the reasons we have chosen it, as we wanted to assess the gap in their performance.

We have conducted several experiments. For MT systems, we compiled lists of stimulus sentences and sequentially translated them into target languages. The LLMs work on different principles, so we needed to create instructions for them. For ChatGPT, we only had to describe the task (in Russian) as “Please translate these sentences into the following language”, and provide it with the lists of sentences and the name of a language; the instruction then was repeated for all of the target languages. Gemini 1.5, however, couldn't cope with the same instruction as it didn't translate all of the sentences one by one and tried to elaborate on the given task instead of doing it, so we had to give it a sentence and a list of target languages at a time. Besides, we encountered a problem with Gemini 1.5 while trying to translate stimuli such as *Vasya ... durak* ‘Vasya … is a fool’: the model has rigid rule-based restrictions which make it plainly refuse to deal with any words considered offensive. That's why we had to replace such stimuli with *Vasya… umnyj* ‘Vasya … is smart’.

In order to evaluate the results of the automatic MT, we used expert translations as a reference; but it must be noted that assessing the quality of translations is not straightforward, so a different translation can still be correct from the point of view of a speaker. We considered different sentences that were close semantically and correct grammatically as ‘correct’ ones. In the table 2, there are also two types of incorrect translations: ‘incorrect’ means that a model chose a wrong operator word (DE DICTO instead of DE RE and vice versa) but the sentence was grammatically well-formed. Sentences with a different meaning were considered translation errors (‘error’). For example, a common issue is for the MT models to translate *samo soboj* ‘naturally’ as ‘by himself’ or ‘alone’, and even ChatGPT makes such mistakes. We also confirm that our chosen LLMs are almost incapable of translating into Ossetian (obviously it is drastically underrepresented in their training data), while Yandex and Google make lots of grammatical mistakes. Surprisingly, ChatGPT shows the worst quality in Ossetian for it hallucinates pseudo-Ossetian sentences with random words.

---

[8] translate.google.com
[9] translate.yandex.com
[10] https://huggingface.co/google/madlad400-3b-mt
[11] https://research.google/blog/recent-advances-in-google-translate/
[12] https://en.wikipedia.org/wiki/Google_Translate
[13] https://www.vellum.ai/llm-leaderboard

		Correct	Incorrect	Error	Total
	Expert	80	10	0	90
	Google	51	1	38	90
	Yandex	71	0	19	90
	ChatGPT	50	5	35	90
Arabic	***Gemini***	71	1	18	90
	Expert	80	10	0	90
	Google	77	9	4	90
	Yandex	69	9	12	90
	ChatGPT	90	0	0	90
Bulgarian	***Gemini***	82	8	0	90
	Expert I	85	5	0	90
	Expert II	80	10	0	90
	Google	80	10	0	90
	Yandex	77	13	0	90
	ChatGPT	85	5	0	90
Danish	***Gemini***	90	0	0	90
	Expert	82	8	0	90
	Google	90	0	0	90
	Yandex	85	5	0	90
	ChatGPT	85	5	0	90
English	***Gemini***	89	1	0	90
	Expert	86	4	0	90
	Google	77	8	5	90
	Yandex	65	11	14	90
	ChatGPT	80	5	5	90
German	***Gemini***	89	1	0	90
	Expert	90	0	0	90
	Google	80	6	4	90
	Yandex	67	7	16	90
	ChatGPT	90	0	0	90
Icelandic	***Gemini***	83	7	0	90
	Expert	80	10	0	90
	Google	0	0	90	90
	Yandex	0	0	90	90
	ChatGPT	0	0	90	90
Ossetian	***Gemini***	0	0	90	90
	Expert	73	17	0	90
	Google	81	9	0	90
	Yandex	77	13	0	90
	ChatGPT	85	5	0	90
Swedish	***Gemini***	90	0	0	90
	Expert	90	0	0	90
	Google	75	0	15	90
	Yandex	81	0	9	90
	ChatGPT	85	0	5	90
Ukrainian	***Gemini***	83	0	7	90

Table 2. The statistics of the results

What our results show from the ‘technical’ point of view, is that 1) Google Translate tends to make less grammatical mistakes than Yandex Translate, although the latter may choose incorrect translations (DE DICTO instead of DE RE and vice versa) less often; 2) ChatGPT translates closer to humans than other models; 3) LLMs surprisingly often choose translation variants close to those of the experts, but not the exact ones, as opposed to the MT systems. 4) Genetic relatedness of languages also matters: the automatic translations for Ukrainian are much closer to the expert ones than for Icelandic or German.

6 Discussion

The results of the study confirm the hypothesis that the contrast of DE RE versus DE DICTO words holds cross-linguistically, despite the class of DE RE words in the target languages is smaller than in Russian: no translator used 7 different DE RE words. The translators generally succeeded in extracting the corrector operator (VER vs AFF) from most DE RE and DE DICTO words, except for the colloquial *konkretno*. Most experts correctly identified the VER operator and placed *konkretno* in the DE RE class. Sporadic deviations, e.g., in the expert translation of *konkretno* via German *bestimmt*^{AFF} or Ossetic *bælwyrdæj*^{AFF} are probably explained by the fact that those experts who treated Rus. *konkretno*^{VER} as a close synonym of Rus. *opredelenno*^{AFF} do not have it in their active vocabulary. The failures of MT systems and LLMs to translate the stimuli with *konkretno*^{VER} can be explained by the insufficient training base, i.e. the lack of the modern input texts containing this discourse item. The same factor was apparently responsible for the mishaps with the stimuli containing the pejorative noun *durak* ‘fool’: the experts rejected the targets like Ger. *Narr*, *Dummkopf*, Ice. *fífl* as dated and weird. This issue however is marginal for the recognition of operators. One expert, a bilingual speaker of Russian and English, refused to provide English equivalents to *i vpryam’*^{VER} and *opredelenno*^{AFF} since he considered these words parasitic. A normative reaction of this kind was not an option for MT and LLMs, while other experts preferred to translate the stimuli even if they did not use them themselves and were not quite sure about their exact meanings.

Most translated versions show a compromise between the horizontal strategy, i. e. the preservation of the same target in Contexts 1 – 5, and the vertical strategy, i. e. the use of the same target standing for different stimuli in contexts of the same type. It is but noteworthy that the choice of the horizontal strategy indicates that the target language is relatively close to Russian regarding the size of DE RE and DE DICTO classes. E. g., our expert in Icelandic provided 5 lexical equivalents for 7 standard Russian DE RE words[14], 9 lexical equivalents for 9 standard Russian DE DICTO words and suggested a correspondence between Rus. *konkretno*^{VER} and Ice. *nefnilega*^{VER}. These moves both confirm the skill of the expert and indicate that Icelandic allows to differentiate nearly as many discourse words expressing the VER | AFF values as Russian. This level was not reached by the MT systems, e. g. Google Translate from Icelandic opted for the vertical translation of 7 Russian DE RE words providing 5 of the them with identical lines but systematically uses different targets in Contexts 1 & 4 versus Contexts 2, 3 & 5. However, the vertical strategy in the hands on an expert tells more about the expert’s attitude to the experiment that about language data. E. g, our expert who approached the vertical strategy in their translations of Russian operator words into German looked for the most idiomatic variants and apparently considered the differentiation of 17 VER | AFF words in the test set of Russian a more futile task than the differentiation of Contexts 1 – 5 containing the VER | AFF insertion. This degree of freedom towards the source data is normally not encouraged in MT systems, which partly explains the gap between expert and MT translation[15]. One expert who consistently used the horizontal strategy for the AFF stimuli developed a special system of translating VER stimuli into English: he translated them lexically or by $\varnothing^{VER}$ but marked the licensing context by a special ‘prefix’.

[14] The stimuli Rus. *dejstvitel’no*^{VER}, *na samom dele*^{VER} and *v samom dele*^{VER} were translated by our expert by identical lines. Meanwhile, the expert managed to preserve the subtle difference between Rus. *dejstvitel’no*^{VER} and *real’no*^{VER} by adding the particle *nú* in the target sentences corresponding to the stimuli with *real’no*^{VER}.
[15] There are hyperparameters like ‘temperature’ in most modern MT systems that allow for a certain degree of freedom, but typically production systems are deterministic by default in order to make their translations consistent and reliable. Also, emerging methods in the area of LLM development may allow and even stimulate variance in translations, as well as in text generation. This may well be seen in our data where Google and Yandex tend to give less variants than LLMs.

(4) Russian ⇒ English

[3.4.] Rus. Vasya **i pravda** ^{VER} oshibsya ⇒ Eng. [$_{\textbf{PREF}}$ **Yes, too bad**]. Vasya **did**^{VER} make a mistake würde. (Expert translation).

[4.5.] Rus. Vasya **na samom dele** ^{VER} ne sobiralsya prixodit'. ⇒ Eng. [$_{\textbf{PREF}}$ **Come to think of it**]: Vasya ∅ ^{VER} didn't even mean to come. (Expert translation).

The 'prefixes' [$_{\textbf{PREF}}$ **Yes, too bad**] and [$_{\textbf{PREF}}$ **Come to think of it**] as such do not convey the meanings of VER or AFF but link the sentences containing VER to a broader context. It is difficult to say whether the corresponding method can be implemented in the MT translation of discourse words. MT depends strongly on the linear context and slight changes in the source sentence may produce seemingly unexplainable differences in the target one, e.g., 'pravda' vs 'i pravda'. There are also issues with the amount of data for the source and target languages: both MT systems and LLMs cope really well with resourced languages like English or German, but their performance deteriorates drastically for Arabic and is virtually null for Ossetian.

7 Conclusions and perspectives

This study confirms the semantic and distributional independence of VER and AFF operators, supporting their universality. LLMs, despite not being specialized for MT tasks, showed remarkable adaptability and context awareness compared to traditional MT systems. The findings highlight the potential of LLMs in nuanced translation tasks and underscore the complexity of translating modal discourse words. Future work will explore custom models and further refine evaluation metrics for translation accuracy.

References

[1] Akinina O. G. Particles from the Class INNA. Functions and Semantics [Chasticy, podobnye INNA. Funkcionirovanie i semantika]. Moscow, 2008.

[2] Averina A. V. German Particles [Chasticy v nemeckom yazyke]. Moscow, 2022.

[3] Baranov A. N., Plungian V. A., Rakhilina E. V. A Guide to Russian Discourse Words [Putevoditel' po diskursivnym slovam russkogo yazyka]. Moscow, 1998.

[4] Brown T., Mann B., Ryder N., Subbiah M., Kaplan J. D., D. Prafulla, A. Neelakantan et al. "Language models are few-shot learners." // *Advances in neural information processing systems 33* (2020), 1877-1901.

[5] Dobrovol'skij D., Pöppel L. Discoursive Construction in the Russian-Swedish Dictionary Database: A Case Study of *v tom-to i N.* // T. Margalitadze, G. Meladze, I. Javakhishvili (eds.), *Proceedings of the XVII Euralex International Congress. Lexicography and Linguistic Diversity*. Tbilisi, 2016, 668 – 677.

[6] Dobrovol'skij D. O., Zaliznjak Anna A. The Parallel Corpus as an Instrument of Semantic Analysis: The German Modal Verb *SOLLEN* [Parallel'nyj korpus kak instrument semanticheskogo analiza: nemeckij modal'nyj glagol *sollen*] // *Proceedings of the International Conference "Corpus Linguistics 2021"*. St-Petersburg, 2021, 209-218.

[7] Frey W. On the categorical status of different dependent clauses. In: J. M. Hartmann & A.Wöllstein (eds.): Propositionale Argumente im Sprachvergleich: Theorie und Empirie. Propositional Arguments in Cross-Linguistic Research: Theoretical and Empirical Issues [Studien zur Deutschen Sprache]. Tübingen: Gunter Narr Verlag, 2023, 364 – 410. DOI: 10.24053/9783823394105

[8] Georgiev P., Ving Ian Lei, Burnell R., Libin Bai, Gulati A., Tanzer G. et al. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530* (2024).

[9] Haspelmath M. Types of Clitics in the World's Languages // *Linguistic Typology at the Crossroads 3-2* (2023), 1 – 59. DOI: 10.6092/issn.2785-0943/16057

[10] Kiseleva K. L., Paillard D. (eds.). Russian Discourse Words: Contextual Variation and Semantic Invariants [Diskursivnye slova russkogo yazyka: Kontekstnoe varjirovanie i semanticheskoe edinstvo]. Moscow, 2003.

[11] Kiseleva K.L. Konechno. Razumeetsya. Estestvenno. In Kiseleva K. L., Paillard D. (eds.) *Russian Discourse Words: Contextual Variation and Semantic Invariants* [Diskursivnye slova russkogo yazyka: Kontekstnoe varjirovanie i semanticheskoe edinstvo] Moscow, 2003.

[12] Kocmi T., Federmann C. Large language models are state-of-the-art evaluators of translation quality. *arXiv preprint arXiv:2302.14520* (2023).

[13] Krifka M. Modelling questions in commitment spaces. In M. Cordes (ed.), *Asking and answering: Rivalling approaches to interrogative methods*, 63–95. Tübingen: Narr Francke Attempto, 2021. DOI: 10.24053/9783823394808

[14] Krifka M. Layers of the clause: Propositions, judgements, commitments, acts. In: J.M. Hartmann & A.Wöllstein (eds.), *Propositionale Argumente im Sprachvergleich: Theorie und Empirie / Propositional Arguments in Cross-Linguistic Research: Theoretical and Empirical Issues* [Studien zur Deutschen Sprache]. Tübingen: Gunter Narr Verlag, 2023, 116 – 183. DOI: 10.24053/9783823394105

[15] Krylova E. V. The Communicative Functions of Danish Modal Particles [Kommunikativnye funkcii modal'nyx chastic v datskom yazyke]. Moscow, 2021.

[16] Levontina I. B. Particles of Speech [Chasticy rechi]. Moscow, 2022.

[17] Lohnstein H. Verum Focus, Sentence Mood, and Contrast. In C. Dimroth and S. Sudhoff (eds.), *The Grammatical Realization of Polarity Contrast: Theoretical, Empirical, and Typological Approaches*. [Linguistik Aktuell/Linguistics Today 249]. 2018, pp. 55–88.

[18] Mikhailov M., Cooper R. Corpus Linguistics for Translation and Contrastive Studies. London and New York: Routledge, 2016. DOI: 10.4324/9781315624570

[19] Mikhailov M. Mind the Source Data! Translation Equivalents and Translation Stimuli from Parallel Corpora. In Vincent X. Wang, Lily Lim, Defeng Li (eds.), *New Frontiers in Translation Studies*. Springer, 2021, 259 — 279. DOI: 10.1007/978-981-16-4918-9

[20] Nelson M. The De Re/De Dicto Distinction. In E.N.Zalta & U.Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2024 Edition), URL = <https:// https://plato.stanford.edu/entries/prop-attitude-reports/dere.html>

[21] Paducheva E. V. Modality [Modal'nost']. In *Data for the Corpus Grammar of Russian. I. The Verb* [Materialy k korpusnoj grammatike russkogo yazyka. I. Glagol]. St.-Petersburg: Nestor-Istoriya, 2016, 19 – 94.

[22] Paducheva E. V. Semantic Studies. The Semantics of Aspect and Tense in Russian. The Semantics of the Narrative [Semanticheskie issledovanija. Semantika vida i vremeni v russkom yazyke; Semantika narrativa]. Moscow, 1996.

[23] Pailliard D. TAKI. In K. L. Kiseleva, D. Paillard (eds.) Russian discourse words. A contextual semantic description [Diskursivnye slova russkogo yazyka. Opyt kontekstno-semanchticheskogo opisaniya]. Moscow, 1998.

[24] Pailliard, Denis. PRAVDA as a discourse word. The guarantee and the perspective [PRAVDA kak diskursivnoe slovo: garant i tochka zrenija]. In Kiseleva K. L., Paillard D. (eds.). *Russian Discourse Words: Contextual Variation and Semantic Invariants* [Diskursivnye slova russkogo yazyka: Kontekstnoe varjirovanie i semanticheskoe edinstvo]. Moscow, 2003, PAGES.

[25] Radford A. Improving language understanding by generative pre-training. June 11, 2018. URL: https://openai.com/index/language-unsupervised/ [Accessed 20.01.2025]

[26] Vaswani A. Attention is all you need. // *Advances in Neural Information Processing Systems, 30*. 2017.

[27] Yanko T. E. The Communicative Status of Expressions Containing the Word PRAVDA [Kommunikativnyj status vyrazhenij so slovom PRAVDA]. In N. D. Arutyunova, N. K. Ryabceva (eds.), *Logical Analysis of the Language. Truth and Veracity in Culture and Language* [Logicheskij analiz yazyka. Istina i istinnost' v kul'ture i yazyke]. Moscow, 1995, 173-177.

[28] Yanko T. E. Communicative Strategies of Russian Speech [Kommunikativnye strategii russkoj rechi]. Moscow, 2001.

[29] Zaliznjak Andrej A. Studies of the Language of Birch Bark Letters [K izucheniju yazyka berestjanyx gramot]. In V. L. Janin, A. A. Zaliznjak. *Novgorodskie gramoty na bereste. Iz raskopok 1984 – 1989 godov* [[Novgorod Birch Bark Letters from the Excavations of 1984–1989]. Moscow, 1993, 191–319.

[30] Zaliznjak Anna A. Russian Language-Specific Words as an Object of Contrastive Analysis [Lingvospecificheskie edinicy russkogo yazyka v svete korpusnogo analiza] // *Computational linguistics and intellectual technologies. Iss. 14 (21).* 2015, 683 – 695.

[31] Zaliznjak Anna A. A Data-Base of Cross-Linguistic Equivalences as an Instrument of Linguistic Analysis [Baza dannyx mezh'yazykovyx ekvivalencij kak instrument lingvisticheskogo analiza] // *Computational linguistics and intellectual technologies. Iss. 15 (22).* 2016, 854 – 866.

[32] Zaliznjak Anna A., Loshchilova E. Y., Kruzhkov M. G., Durnovo A. A., Buntman N. V., Egorova A. Y., Zatsman I. M. The Library of Discourse Words Annotations in the Parallel Russian-French Texts, version 2.0 [Biblioteka annotacij diskursivnyx slov v russko-francuzskix parallel'nyx tekstax (versiya 2.0). Certificate of registration of the database RU 2022623727, 27.12.2022.

[33] Zimmerling A. Biased Questions and Hamblin Semantics // *Typology of Morphosyntactic Parameters*. 2023. Vol. 6. Iss. 2, 92-135. DOI: 10.37632/PI.2023.94.43.005

[34] Zimmerling A. Really: Syntactics without Semiotics? [Konkretno: sintaktika bez semiotiki?] // *Slovo. Ru: Baltic Accent*. 2023. № 3, 125-153. DOI: 10.5922/2225-5346-2023-3-9

[35] Zimmerling A. From God's Truth to Concession [Ot Bozhjej pravdy k ustupke]. In M. S. Milovanova (ed.), *Axiology and Cognitive Linguistics* [Aksiologiya i kognitivnaya lingvistika]. Moscow, 2024, 346 – 353. DOI: 10.5281/zenodo.15098320

[36] Zimmerling A.V., Yanko T. E. Verification and Truth. From Predicate to Connector [Verifikacija i pravda. Ot predikata k konnektoru]. In I. M. Kobozeva, N. V. Serdobolskaya, A. I. Kryukova (eds.), *Connectors in Modern Russian. Semantics, Syntax, Pragmatics, Diachrony* [Konnektory sovremennogo russkogo yazyka: semantika i sintaksis, pragmatika i diahroniya. Moscow, 2025.